Optimization methods and stability of inclusions in Banach Spaces

Klatte, Diethard ; Kummer, Bernd

Abstract: Our paper deals with the interrelation of optimization methods and Lipschitz stability of multifunctions in arbitrary Banach spaces. Roughly speaking, we show that linear convergence of several first order methods and Lipschitz stability mean the same. Particularly, we characterize calmness and the Aubin property by uniformly (with respect to certain starting points) linear convergence of descent methods and approximate projection methods. So we obtain, e.g., solution methods (for solving equations or variational problems) which require calmness only. The relations of these methods to several known basic algorithms are discussed, and errors in the subroutines as well as deformations of the given mappings are permitted. We also recall how such deformations are related to standard algorithms like barrier, penalty or regularization methods in optimization.

DOI: https://doi.org/10.1007/s10107-007-0174-9

Originally published at:
Klatte, Diethard; Kummer, Bernd (2009). Optimization methods and stability of inclusions in Banach Spaces. Mathematical Programming, 117(1-2):305-330.
DOI: https://doi.org/10.1007/s10107-007-0174-9

Math. Program., Ser. B (2009) 117:305–330
DOI 10.1007/s10107-007-0174-9

FULL LENGTH PAPER

Optimization methods and stability of inclusions in Banach spaces

Diethard Klatte · Bernd Kummer

Received: 12 June 2006 / Accepted: 31 August 2006 / Published online: 20 July 2007

Abstract Our paper deals with the interrelation of optimization methods and Lipschitz stability of multifunctions in arbitrary Banach spaces. Roughly speaking, we show that linear convergence of several first order methods and Lipschitz stability mean the same. Particularly, we characterize calmness and the Aubin property by uniformly (with respect to certain starting points) linear convergence of descent methods and approximate projection methods. So we obtain, e.g., solution methods (for solving equations or variational problems) which require calmness only. The relations of these methods to several known basic algorithms are discussed, and errors in the subroutines as well as deformations of the given mappings are permitted. We also recall how such deformations are related to standard algorithms like barrier, penalty or regularization methods in optimization.

Keywords Generalized equation · Variational inequality · Perturbation · Regularization · Stability criteria · Metric regularity · Calmness · Approximate projections · Penalization · Successive approximation · Newton's method

Mathematics Subject Classification (2000) 49J52 · 49K40 · 90C31 · 65Y20

This paper is dedicated to Professor Stephen M. Robinson on the occasion of his 65th birthday.

D. Klatte (✉)
Institut für Operations Research, Universität Zürich, Moussonstrasse 15,
CH-8044 Zürich, Switzerland
e-mail: klatte@ior.uzh.ch

B. Kummer
Institut für Mathematik, Humboldt–Universität zu Berlin, Unter den Linden 6,
10099 Berlin, Germany
e-mail: kummer@mathematik.hu-berlin.de

1 Introduction

It is well-known that, in the context of various solution methods, statements on "stability" of the equation are helpful tools for verifying convergence.

In this paper, we show that the applicability of certain solution methods is even equivalent to some classical types of stability for equations and inclusions (also called generalized equations) as well. In other words, we present solution procedures which converge (locally and with linear order of convergence) exactly under the mentioned stability condition and present stability criteria in terms of such solution procedures. So we hope that our approach helps to decrease the gap between stability and its main applications, the behavior of solution methods.

Our basic model is the generalized equation

$$\text{Find } x \text{ such that } p \in F(x), \qquad F : X \rightrightarrows P, \tag{1.1}$$

where $p \in P$ is a canonical parameter, P, X are Banach spaces and F is a closed multifunction, i.e., $F(x) \subset P$ and the graph of F, $\operatorname{gph} F = \{(x, p) \mid p \in F(x)\}$, is a closed set.

System (1.1) describes solutions of equations as well as stationary or critical points of various variational conditions. It was Stephen M. Robinson who introduced in several basic papers [32–34] generalized equations as a unified framework for mathematical programs, complementarity problems and related variational problems. His work influenced much the development of stability analysis and of foundations of solution methods in the last 20–25 years, for a survey of these developments see [35].

In particular, for optimization problems, a deep analysis of critical points is mainly required in hierarchic models which arise as "multiphase problems" if solutions of some or several problems are involved in a next one. For various concrete models and solution methods we refer, e.g. to [8,13,30], while a big scope of continuity results for critical values and solutions of optimization problems in $\mathbb{R}^n$ can be found in [2]. Several further applications of model (1.1) are known for optimization problems, for describing equilibria and other solutions in games, in so-called MPECs and stochastic and/or multilevel models. We refer, e.g. to [1,3,6,8,13,21,22,30,36] for the related settings.

We will study *local* stability of solutions to (1.1), i.e., we consider the map $S(p) = F^{-1}(p)$ near some particular solution $x^0 \in S(p^0)$.

As already in [23], we intend to characterize stable behavior of the solutions by means of usual analytical techniques and by the behavior (uniform linear convergence for starting points near (x^0, p^0)) of methods for solving (1.1) in original form or under additional "small" nonlinear perturbations like

$$p \in h(x) + F(x), \quad h : X \to P, \tag{1.2}$$

where "+" denotes the elementwise sum.

Here, in contrast to [23], we permit errors in the iteration schemes. This is essential since it allows us to consider arbitrary Banach spaces X and P and to avoid preparations via Ekeland's variational principle [12]. The latter can be done since we shall not aim at using the close relations between stability and injectivity of certain generalized derivatives (which do not hold in general Banach spaces). For approaches studying these relations, we refer the reader to the monographs [1,6,13,21,29,36]. Notice, however, that (up to now) there is no derivative-criterion for the Aubin property or calmness of Lipschitz functions in arbitrary Banach spaces (even less for multifunctions). In view of calmness, our discussion after Theorem 3 (see the torus-argument) explains one reason for this fact. Furthermore, the way from derivative characterizations of "stability" to solution methods (particularly in Banach space) is usually long and restricted to special problem-classes only. We shall establish a general and direct approach.

For showing and characterizing the Aubin property, particular methods (basically of Newton-type and successive approximation) have been already exploited in several papers, cf. [7,10,15,21,24,25,28]. Further algorithmic approaches for verifying stability of intersections of multifunctions, can be found in [17] and [26]. In [17], calmness has been verified via Newton's method for semismooth functions. In [26], the Aubin property has been characterized by MFCQ-like conditions in B-spaces.

Notice however, that Newton-type methods cannot be applied in our context due to lack of differentiability (or of "semi-smoothness"), and successive approximation techniques fail to work under calmness alone. Also the proper projection and penalty methods applied in [23] require additional hypotheses for the existence of solutions in Banach spaces.

The paper is organized as follows. In Sect. 2, some notions of local Lipschitz stability are introduced which are well-known from the literature (cf. e.g. [1,3,13,21,36]), we compile crucial interrelations between them and we point out the differences between known conditions of calmness and Aubin property for usual C^1 constraints in finite dimension.

The main Sect. 3 is devoted to general stability criteria in terms of solution procedures. After starting with some basic algorithmic scheme ALG1 (which may be seen as a descent method), Theorem 2 shows that linear convergence of an approximate projection method PRO(γ) for computing some $x_\pi \in S(\pi)$ plays a key role. In this way, we characterize calmness and the Aubin property in a constructive manner and indicate the difference between both stability properties in an algorithmic framework.

In particular, we pay attention to the case of F being a locally Lipschitz operator and characterize calmness (Theorem 3, 4 via ALG2, ALG3) for (finite or infinite) systems of inequalities. Using ALG3, we solve linear inequalities (with a convex norm-condition) in order to characterize calmness for a system of nonconvex C^1-inequalities, or in order to solve this nonconvex system under calmness.

In Sect. 4, we discuss further interpretations of ALG1 and PRO(γ) via projections (e.g. Feijer method) and penalizations as well as relations to modified successive approximation and to Newton methods.

Finally, Sect. 5 is reserved for discussing the algorithms for nonlinearly perturbed inclusions. In particular, modified successive approximation is used for verifying the Aubin property (and computing related solutions) of the system (1.2)

2 Notions of local Lipschitz stability

In the whole paper, $S : P \rightrightarrows X$ is a closed multifunction (the inverse of F), P, X are Banach spaces and $z^0 = (p^0, x^0) \in \operatorname{gph} S$ is a given point. We write ζ^0 in place of (x^0, p^0) and say that some property holds *near* x if it holds for all points in some neighborhood of x. Further, let B denote the closed unit ball in the related space and

$$S_\varepsilon(p) := S(p) \cap (x^0 + \varepsilon B) := S(p) \cap \{x \mid d(x, x^0) \le \varepsilon\}.$$

Note that we often write $d(x, x^0)$ for the (induced) distance in X, for better distinguishing terms in the spaces P and X (moreover, often X may be a complete metric space). By conv M we denote the convex hull of a set M.

The following definitions generalize typical local properties of the multivalued inverse $S = f^{-1}$ or of level sets $S(p) = \{x \mid f(x) \le p\}$ for functions $f : M \subset X \to \mathbb{R}$.

Definition 1 Let $z^0 = (p^0, x^0) \in \operatorname{gph} S$.

a. S is said to be *pseudo–Lipschitz* or to have the *Aubin property* at z^0 if

$$\exists\, \varepsilon, \delta, L > 0 \text{ such that } S_\varepsilon(p) \subset S(p') + L\|p' - p\| B \;\forall p, p' \in p^0 + \delta B. \tag{2.1}$$

b. If for sufficiently small ε and $\|p - p^0\|$, $S_\varepsilon(p)$ is even a singleton in (2.1), we call S strongly Lipschitz *stable* (s.L.s.) at (p^0, x^0).

c. S is said to be *calm* at z^0 if (2.1) holds for $p' = p^0$, i.e.,

$$\exists\, \varepsilon, \delta, L > 0 \text{ such that } S_\varepsilon(p) \subset S(p^0) + L\|p - p^0\| B \;\forall p \in p^0 + \delta B. \tag{2.2}$$

d. S is said to be *locally upper Lipschitz* (locally u.L.) at z^0 if

$$\exists\, \varepsilon, \delta, L > 0 \text{ such that } S_\varepsilon(p) \subset x^0 + L\|p - p^0\| B \;\;\forall p \in p^0 + \delta B. \tag{2.3}$$

e. S is said to be *lower Lipschitz* or *Lipschitz lower semicontinuous* (Lipschitz l.s.c.) at z^0 if

$$\exists\, \delta, L > 0 \text{ such that } S(p) \cap (x^0 + L\|p - p^0\| B) \ne \emptyset \;\forall p \in p^0 + \delta B. \tag{2.4}$$

Remark 1 Let us add some comments concerning the notions just defined.

(i) The constant L is called a *rank* of the related stability.

(ii) If $S = f^{-1}$ is the inverse of a C^1 function $f : \mathbb{R}^n \to \mathbb{R}^n$ with $S(p^0) = \{x^0\}$, all these properties coincide and are equivalent to det $Df(x^0) \ne 0$. If f is only locally Lipschitz, even more for model (1.1), they are quite different.

(iii) With respect to the Aubin property (2.1) it is equivalent to say that S^{-1} is metrically regular resp. pseudo-regular, see e.g. [21] for details. Strong Lipschitz stability of S is the counterpart of strong regularity of S^{-1} as used in [21]. Note that "strong regularity" of multifunctions has been also defined in an alternative manner in [33] via local linearizations and requiring that the linearized map is s.L.s. in the above sense.

(iv) Setting $p = p^0$ in (2.1), one obtains $S(p') \neq \emptyset$ for $p' \in p^0 + \delta B$ due to $x^0 \in S_\varepsilon(p^0)$. Thus $p^0 \in \operatorname{int}\operatorname{dom} S_\varepsilon$ follows from (2.1).
This inclusion means that solutions to (1.1) are locally *persistent*, and the Lipschitz l.s.c. property quantifies this persistence in a Lipschitzian manner.

(v) The Aubin property is persistent with respect to small variations of $z^0 \in \operatorname{gph} S$ since (2.1) holds (if at all) also for $L, \varepsilon' = \frac{\varepsilon}{2}, \delta' = \frac{\delta}{2}$ and $z^{0'} = (p^{0'}, x^{0'}) \in \operatorname{gph} S$ with $d(x^{0'}, x^0) < \varepsilon'$ and $\|p^{0'} - p^0\| < \delta'$. Decreasing $\|p^{0'} - p^0\|$ if necessary, one obtains the same for the strong Lipschitz stability. On the contrary, the properties c., d. and e. in Definition 1 may fail to hold after arbitrarily small variations of $z^0 \in \operatorname{gph} S$.

Remark 2 For fixed $z^0 = (p^0, x^0) \in \operatorname{gph} S$, one easily sees by the definitions:

(i) S is locally u.L. at z^0 $\Leftrightarrow$ S is calm at z^0 and x^0 is isolated in $S(p^0)$.

(ii) S is pseudo-Lipschitz at z^0 $\Leftrightarrow$ S is Lipschitz l.s.c. at all points $z \in \operatorname{gph} S$ near z^0 with fixed constants δ and L.

(iii) S is pseudo-Lipschitz at z^0 $\Leftrightarrow$ S is both calm at all $z \in \operatorname{gph} S$ near z^0 with fixed constants ε, δ, L and Lipschitz l.s.c. at z^0.

The example of C^1 constraints in $\mathbb{R}^n$

For every constraint system of a usual optimization model in $X = \mathbb{R}^n$, namely

$$\Sigma(p_1, p_2) = \{x \in \mathbb{R}^n \mid g(x) \leq p_1,\ h(x) = p_2\}, \quad (g, h) \in C^1(\mathbb{R}^n, \mathbb{R}^{m_1+m_2}), \tag{2.5}$$

the Aubin property can be characterized by elementary and intrinsic means. Let $z^0 = (0, x^0) \in \operatorname{gph} \Sigma$.

Lemma 1 *For the multifunction Σ (2.5), the following statements are equivalent*:

1. *Σ is Lipschitz l.s.c. at z^0.*
2. *Σ obeys the Aubin property at z^0.*
3. *The Mangasarian-Fromowitz constraint qualification* (*MFCQ*) *holds at z^0, i.e.,*

$$\begin{aligned} &\operatorname{rank} Dh(x^0) = m_2, \textit{ and } \exists u \in \ker Dh(x^0) \\ &\textit{such that } g_i(x^0) + Dg_i(x^0)u < 0\ \forall i. \end{aligned} \tag{2.6}$$

Proof The equivalence of 2. and 3. is well-known, it follows from Robinson's basic paper [31], by taking the equivalence of Aubin property and metric regularity into account. Further, 2. implies 1., by Remark 2 (ii).

The remaining implication 1. $\Rightarrow$ 3. is a consequence of $g, h \in C^1$: Since, for small $\|p\|$, solutions $x(p) \in \Sigma(p)$ exist with $\|x(p) - x^0\| \leq L\|p\|$, one obtains first $\operatorname{rank} Dh(x^0) = m_2$ (otherwise choose $p(t) = (0, tp_2)$ where $p_2 \notin \operatorname{Im} Dh(x^0)$, $t \downarrow 0$) and next the second condition in (2.6) by considering $p(t) = (tp_1, 0)$, where $p_1 = (-1, \ldots, -1)$, and choosing a cluster point u of $\|x(p(t)) - x^0\|/t$. ◻

Analyzing calmness of Σ at z^0 seems to be even simpler since it suffices to investigate calmness of the inequality system

$$\widetilde{\Sigma}(t) = \{x \in \mathbb{R}^n \mid g_i(x) \le t,\ -t \le h_k(x) \le t \quad \forall\, i = 1, \dots, m_1,\ k = 1, \dots, m_2\} \tag{2.7}$$

at $(0, x^0) \in \mathbb{R} \times X$ only and, in addition, calmness requires less than the Aubin property. Nevertheless, its characterization is more complicated, provided the functions involved are not piecewise linear (then calmness holds true). So it is known from [18] that the Abadie constraint qualification required at $x^0 \in M = \Sigma(0)$, is necessary (but not sufficient) for calmness of Σ at $(0, x^0)$. Furthermore, there are several sufficient calmness conditions which fit to our problem class (2.5), see e.g. [17,18]. For example, Theorem 3 of [18] says for Σ (2.5) without equations:

> *Σ is calm at $(0, x^0) \in \operatorname{gph} \Sigma$ if (at x^0) both the Abadie CQ holds true and MFCQ with respect to the set $M(J) := \{x \mid g_i(x) \le 0\ \forall i \in J\}$ is satisfied, whenever J fulfills $g_i(\xi^k) = 0$ ($\forall i \in J$, $\forall k \in \mathbb{N}$) for $\xi^k \to x^0$ with $\xi^k \in \operatorname{bd} M \setminus \{x^0\}$.*

This sufficient condition is not satisfied for the linear (and calm) example

> $\Sigma(p_1, p_2) = \{(x_1, x_2) \mid x_2 \le p_1,\ -x_2 \le p_2\}$: MFCQ does not hold at $0 \in M(\{1, 2\})$.

Surprisingly, we nowhere found a necessary and sufficient calmness criterion in terms of the original data, though for this situation there is a condition which is similar to MFCQ, cf. Theorem 3.

For *convex* C^1 inequalities, calmness of Σ at $(0, x^0)$ holds true if and only if the Abadie CQ holds at all points of $\Sigma(0)$ in some neighborhood of x^0, see [5,27]. However, checking the latter is nontrivial, too (since—up to now—there is no efficient analytical condition for this property).

3 Stability and algorithms

Let $S = F^{-1} : P \rightrightarrows X$ be given as under (1.1). Though we are speaking about closed *multifunctions* which act between Banach spaces, our stability properties for S are classical properties of *non-expansive, real-valued functions* only.

This is true since calmness at (p^0, x^0) is a *monotonicity property* with respect to two canonically assigned *Lipschitz functions*: the distance $\psi(x, p) = \operatorname{dist}((p, x), \operatorname{gph} S)$ and the distance of x to $S(p^0)$. In terms of ψ, *calmness* of S at $(p^0, x^0) \in \operatorname{gph} S$ equivalently means that

$$\exists \varepsilon > 0,\ \alpha > 0 \text{ such that } \alpha \operatorname{dist}(x, S(p^0)) \le \psi(x, p^0) \quad \forall x \in x^0 + \varepsilon B, \tag{3.1}$$

where ψ is defined via the norm $\|(p, x)\| = \max\{\|p\|, \|x\|\}$ or some equivalent norm in $P \times X$. For details of the equivalence proof and estimates of ψ for particular systems, we refer to [20]. Condition (3.1) requires that $\psi(., p^0)$ increases in a Lipschitzian manner if x leaves $S(p^0)$. Clearly, this property depends on the local structure of the boundaries of $\operatorname{gph} S$ and $S(p^0)$ and (approximate) normal directions only. For convex multifunctions (i.e. gph S is convex), ψ and $d(., S(p^0))$ are even convex and (globally) Lipschitz.

Combined with Remark 2 (iii), condition (3.1) characterizes the Aubin property, too. Concerning similar characterizations of other stability properties we refer to [23].

The distance ψ can be also applied for both *characterizing optimality and computing* solutions in optimization models via penalization [20,26] and [21, Chapt. 2]; for the particular context of exact penalization techniques, see also [4,6,9].

The approximate minimization of ψ (defined by a norm $\|(p, x)\| = \lambda^{-1}\|p\| + \|x\|$) will play a main role below.

3.1 The algorithmic framework

We continue considering closed mappings $S = F^{-1}$.

Given $(p, x) \in \text{gph}\, S$ near $z^0 = (p^0, x^0)$ and π near p^0 (*briefly: given initial points* x, p, π *near* z^0), we want to determine some $x_\pi \in S(\pi)$ with $d(x_\pi, x) \le L\|\pi - p\|$ *by algorithms.* The existence of x_π is claimed under the Aubin property (or under calmness if $\pi = p^0$).

Notice that, under the viewpoint of solution methods, we usually have $\pi = p^0 = 0$, and $p^0 \in F(.)$ is the "equation" we want to solve with start at some $(x, p) \in \text{gph}\, F$.

In stability theory, some solution $x^0 \in S(p^0)$ is considered to be given and the local behavior of solutions to $\pi \in F(.)$ (π near p^0) is of interest.

So we unify these two viewpoints by discussing how $\pi \in F(.)$ can be solved (and solutions can be estimated) with initial points (p, x) near z^0. Evidently, it suffices to minimize $d(\xi, x)$ s.t. $\xi \in F^{-1}(\pi)$ for this purpose. However, this nonlinear problem requires some concrete algorithm, in general, and the existence of a minimizer is questionable, too. Therefore, we are interested in procedures which find x_π with well-defined rate of convergence, *exactly* under the Aubin property (or under calmness, respectively).

By saying that some algorithm has this specific property (for initial points near z^0) we try to connect stability and solution methods in a direct and fruitful manner.

Due to the aimed generality, our crucial methods ALG1 and PRO(γ) are of quite simple type. Nevertheless they involve several more or less fast local methods under additional assumptions.

The subsequent first algorithm should be understood like a framework for more concrete procedures which compute $x_\pi \in S(\pi)$. Suppose that some $\lambda \in (0, 1)$ is given.

ALG1 Put $(p^1, x^1) = (p, x) \in \text{gph}\, S$ and choose $(p^{k+1}, x^{k+1}) \in \text{gph}\, S$ in such a way that

$$\begin{array}{llcll} \text{(i)} & \|p^{k+1} - \pi\| - \|p^k - \pi\| & \le & -\lambda\, d(x^{k+1}, x^k) & \text{and} \\ \text{(ii)} & \|p^{k+1} - \pi\| - \|p^k - \pi\| & \le & -\lambda\, \|p^k - \pi\|. & \end{array} \tag{3.2}$$

Definition 2 We call ALG1 *applicable* if related (p^{k+1}, x^{k+1}) exist in each step (for some fixed $\lambda > 0$).

Having calmness in mind, we apply the same algorithm with fixed $\pi \equiv p^0$.

Interpretation:

Identifying p^k with some element $f(x^k) \in F(x^k)$ condition (3.2)(i) requires more familiar

$$\frac{\|f(x^{k+1}) - \pi\| - \|f(x^k) - \pi\|}{d(x^{k+1}, x^k)} \le -\lambda \quad \text{for } x^{k+1} \ne x^k, \tag{3.3}$$

and (3.2)(ii) is one of various conditions which ensure $\|f(x^k) - \pi\| \to 0$ for this (non-increasing) sequence. In this interpretation, ALG1 is a *descent method* for the function $x \mapsto \|f(x) - \pi\|$.

Reducing the stepsize:

As in every method of this type, one may start with some $\lambda = \lambda_1 > 0$ and, if (p^{k+1}, x^{k+1}) satisfying (3.2) cannot be found, decrease λ by a constant factor, e.g. $\lambda_{k+1} = \frac{1}{2}\lambda_k$ while $(p^{k+1}, x^{k+1}) := (p^k, x^k)$ remains unchanged. In this form, *being applicable* coincides with

$$\inf \lambda_k \ge \alpha > 0,$$

and we shall need the same α with respect to the possible starting points.

This modification or reduction of λ (*like for the Armijo-Goldstein stepsize rule in free minimization problems*) *is possible for all algorithms we shall speak about, though we make explicitly use of it only for ALG2 and ALG3, cf. Theorem 4.*

Theorem 1 *Let $S : P \rightrightarrows X$ be closed. If ALG1 is applicable for given initial points x, p, π near z^0, then the sequence converges $(p^k, x^k) \to (\pi, x_\pi) \in \operatorname{gph} S$, and*

$$d(x_\pi, x) \le \frac{1}{\lambda}\|\pi - p\|. \tag{3.4}$$

Moreover,

(i) *The Aubin property of S holds at $z^0 = (p^0, x^0) \Leftrightarrow$ ALG1 is applicable, for some fixed $\lambda \in (0, 1)$ and all initial points x, p, π near z^0.*
(ii) *The same statement, however with fixed $\pi \equiv p^0$, holds in view of calmness of S at z^0.*

Proof If ALG1 is applicable then, beginning with $n = 1$ and $x^1 = x$, the estimate

$$d(x^{n+1}, x) \le \sum_{k=1}^{n} d(x^{k+1}, x^k) \le \frac{\|p^1 - \pi\| - \|p^{n+1} - \pi\|}{\lambda} \tag{3.5}$$

follows from (3.2)(i) by complete induction. So, a Cauchy sequence $\{x^k\}$ will be generated and (3.5) ensures (3.4) for the limit $x_\pi = \lim x^k$. Taking also (3.2)(ii) into account, it follows $p^k \to \pi$. Since S is closed, so also $x_\pi \in S(\pi)$ is valid.

(i), (ii) ($\Rightarrow$) Let the Aubin property be satisfied with related constants L, ε, δ in (2.1). Then we obtain the existence of the next iterates whenever $0 < \lambda < L^{-1}$ and $\|\pi - p^0\| + d((p,x), z^0)$ was small enough. Indeed, if $\hat{\varepsilon} := \min\{\varepsilon, \delta\}$ and

$$\max\{ \frac{\|p - p^0\| + \|\pi - p^0\|}{\lambda} , d(x, x^0) \} < \tfrac{1}{2}\hat{\varepsilon}$$

then $d(x^k, x^0) < \hat{\varepsilon}$ and $\|p^k - p^0\| < \hat{\varepsilon}$ follow from (3.5) by induction. Thus, for any p^{k+1} in the convex hull conv $\{p^k, \pi\}$ satisfying (3.2)(ii) there is some $x^{k+1} \in S(p^{k+1})$ such that

$$d(x^{k+1}, x^k) \le L\|p^{k+1} - p^k\| \le \frac{\|p^{k+1} - p^k\|}{\lambda} = \frac{\|p^k - \pi\| - \|p^{k+1} - \pi\|}{\lambda}.$$

Hence also x^{k+1} exists as required in (3.2)(i).

Having only calmness, the existence of a related element $x^{k+1} \in S(p^{k+1})$ is ensured by setting $p^{k+1} = \pi = p^0$ (whereafter the sequence becomes constant).

(i), (ii) ($\Leftarrow$) If the Aubin property is violated and $\lambda > 0$, then (by definition) one finds points $(p, x) \in \text{gph}\, S$ arbitrarily close to z^0, and π arbitrarily close to p^0, such that dist $(x, S(\pi)) > \frac{\|p-\pi\|}{\lambda}$. Consequently, it is also impossible to find some related x_π by ALG1.

In view of calmness, the same arguments apply to $\pi \equiv p^0$. □

Remark 3 (i) Theorem 1 still holds after replacing (3.2)(ii) by any condition which ensures, along with (3.2)(i), that $p^k \to \pi$. Hence, instead of (3.2)(ii), one can require that the stepsize is linearly bounded below by the current error

$$d(x^{k+1}, x^k) \ge c\, \|p^k - \pi\| \text{ for some } c > 0. \tag{3.6}$$

Evidently, (3.2)(i) and (3.6) imply (3.2) with new λ.

(ii) Generally, (3.6) does not follow from (3.2), take the function $F(x) = \sqrt[3]{x}$. So requiring (3.2) is weaker than (3.2)(i) and (3.6).

(iii) Theorem 1 remains true (with the same proof) if one additionally requires $p^k \in \text{conv}\,\{p^1, \pi\} \forall k$ in (3.2).

Without considering sequences explicitly, the statements (i), (ii) of Theorem 1 can be written as stability criterions.

Corollary 1 (i) *The Aubin property of* S *holds at* $z^0 = (p^0, x^0) \Leftrightarrow$ *For some* $\lambda \in (0, 1)$ *and all initial points* x, p, π *near* z^0 *there exists some* $(p', x') \in \text{gph}\, S$ *such that*

$$\begin{array}{llll} (i) & \|p' - \pi\| - \|p - \pi\| & \le & -\lambda\, d(x', x) \quad \textit{and} \\ (ii) & \|p' - \pi\| - \|p - \pi\| & \le & -\lambda\, \|p - \pi\|. \end{array} \tag{3.7}$$

(ii) *The same statement, with fixed* $\pi \equiv p^0$*, holds in view of calmness of* S *at* z^0.

Proof It suffices to show that ALG1 is applicable under (3.7). Denoting (p', x') by $\phi(p, x)$, define

$$(p^1, x^1) = (p, x) \text{ and } (p^{k+1}, x^{k+1}) = \phi(p^k, x^k). \tag{3.8}$$

Due to (3.5), (p^n, x^n) belongs to an arbitrary small neighborhood Ω of z^0 for all initial points $(x, p), \pi$ sufficiently close to z^0 and p^0, respectively. Hence ALG1 is applicable. □

3.2 The behavior of ALG1

The similarity of the statements for calmness and the Aubin property does not imply that ALG1 runs in the same way under each of these properties:

Aubin property:

If ALG1 is applicable for all initial points near $z^0 \in \operatorname{gph} S$, we can first fix any $p^{k+1} \in \operatorname{conv}\{p^k, \pi\}$ satisfying (3.2)(ii) and next find (since the Aubin property holds at z^0 by Theorem 1 and (p^k, x^k) is close to z^0) some $x^{k+1} \in S(p^{k+1})$ satisfying (3.2)(i).

In other words, x_π can be determined by small steps. This is not important for estimating $d(x, x_\pi)$, but for constructing concrete algorithms which use local information for F near (p^k, x^k) in order to find (p^{k+1}, x^{k+1}).

Calmness:

Though every sequence in (3.2) leads us to $x_\pi \in S(\pi)$, we can guarantee that some feasible x^{k+1} exists *for some already given* p^{k+1}, only if $p^{k+1} = \pi = p^0$.

In other words, the sequence could be trivial, $(p^k, x^k) = (\pi, x_\pi)\ \forall\, k \geq k_0$, since calmness allows (by definition) that $S(p) = \emptyset$ for $p \notin \{p^1, p^0\}$. In this case, local information for F near (p^k, x^k) cannot help to find x^{k+1} for given $p^{k+1} \in \operatorname{int}\operatorname{conv}\{p^1, \pi\}$.

However, for many mappings which describe constraint systems or solutions of variational inequalities, this is not the typical situation. In particular if gph S is convex then $S(p^{k+1}) \neq \emptyset$ holds for each $p^{k+1} \in \operatorname{conv}\{p^1, \pi\}$ (since $S(\pi)$ and $S(p^1)$ are non-empty by assumption). This remains true if gph S is (as in various MPCP problems) a finite union of closed convex sets C_i since $I(z) := \{i \mid z \in C_i\} \subset I(z^0)$ holds for all initial points $z = (p, x) \in \operatorname{gph} S$ near z^0. More general, it would be sufficient that the sets $F(x_\pi + \varepsilon B)$ are star-shaped with center π.

3.3 Stability in terms of approximate projections

The following *approximate projection method* (onto gph S) has, in contrast to ALG1, the advantage that iteration points throughout exist (for $\gamma > 0$). “Stability” is now characterized by linear order of convergence. Let $\gamma \geq 0$.

PRO(γ) Put $(p^1, x^1) = (p, x) \in \operatorname{gph} S$ and choose $(p^{k+1}, x^{k+1}) \in \operatorname{gph} S$ in such a way that

$$d(x^{k+1}, x^k) + \frac{\|p^{k+1} - \pi\|}{\lambda} \leq \inf_{(p', x') \in \operatorname{gph} S} \left[d(x', x^k) + \frac{\|p' - \pi\|}{\lambda} \right] + \gamma \|p^k - \pi\|. \tag{3.9}$$

Theorem 2 (i) *The Aubin property of S holds at $z^0 = (p^0, x^0) \Leftrightarrow$ PRO(γ) generates, for some $\lambda > 0$ and all initial points x, p, π near z^0, a sequence satisfying*

$$\lambda\, d(x^{k+1}, x^k) + \|p^{k+1} - \pi\| \leq \theta \|p^k - \pi\| \quad \textit{with some fixed } \theta < 1. \tag{3.10}$$

(ii) *The same statement, with $\pi \equiv p^0$, holds in view of calmness of S at z^0.*

Note. Obviously (3.10) means

$$\|p^{k+1} - \pi\| - \|p^k - \pi\| \leq -\lambda\, d(x^{k+1}, x^k) - (1 - \theta)\|p^k - \pi\|$$

which implies (3.2)(i) and again convergence $x^k \to x_\pi \in S(\pi)$ satisfying (3.4). Further, having the related stability property, the next proof shall indicate that one may apply PRO(γ) with any positive γ, provided that λ is sufficiently small, see the requirement $\lambda(L + \gamma) < 1$.

Proof (i) ($\Rightarrow$) Suppose the Aubin property with rank L, and fix $\lambda \in (0,\ (L + \gamma)^{-1})$. Considering again points near (p^0, x^0) one may apply the existence of $\hat{x} \in S(\pi)$ with $d(\hat{x}, x^k) \leq L\|\pi - p^k\|$. This yields for the approximate minimizer in (3.9)

$$\begin{aligned} d(x^{k+1}, x^k) + \frac{1}{\lambda}\|p^{k+1} - \pi\| &\leq d(\hat{x}, x^k) + \frac{1}{\lambda}\|\pi - \pi\| + \gamma\|p^k - \pi\| \\ &\leq (L + \gamma)\|p^k - \pi\| \end{aligned}$$

and implies

$$\lambda\, d(x^{k+1}, x^k) + \|p^{k+1} - \pi\| \leq \lambda\, (L + \gamma)\, \|p^k - \pi\|$$

as well as (3.10) with $\theta = \lambda(L + \gamma) < 1$.

($\Leftarrow$) Conversely, assume that PRO(γ) (or any algorithm) generates a sequence satisfying (3.10) with some $\lambda > 0$, $\theta \in (0, 1)$ and all related initial points. Then also (3.2)(i) is valid for the current sequences and $\|p^{k+1} - \pi\|$ vanishes. By Theorem 1 and Remark 3(i) so the Aubin property must be satisfied.

(ii) Applying the modification for calmness in the same manner, the assertion follows. □

Combining condition (3.1) for calmness of S at z^0 (with the norm $\lambda\|.\|_X + \|.\|_P$ in $X \times P$) and condition (3.10) with $\pi = p^0$, one directly obtains the calmness estimate

$$\theta\, \|p^k - p^0\| \geq \lambda\, d(x^{k+1}, x^k) + \|p^{k+1} - \pi\| \geq \alpha\, \operatorname{dist}(x^k, S(p^0)). \tag{3.11}$$

3.4 The particular case of $F = f$ being a locally Lipschitz operator

We shall see that, in this situation, condition (3.2) can be written (up to a possibly new constant λ) as

$$\|f(x^{k+1})-\pi\|-\|f(x^k)-\pi\| \le -\lambda\, d(x^{k+1},x^k) \text{ and } d(x^{k+1},x^k) \ge \lambda\|f(x^k)-\pi\| \tag{3.12}$$

or equivalently as

$$\|f(x^k)-\pi\|-\|f(x^{k+1})-\pi\| \ge \lambda d(x^{k+1},x^k) \ge \lambda^2\|f(x^k)-\pi\|.$$

This permits a stability characterizations in terms of minimizing sequences with a stepsize estimate as in Remark 3(i).

Corollary 2 *Let $f : X \to P$ be locally Lipschitz near a zero x^0. Then $S = f^{-1}$ obeys the Aubin property at $(0, x^0)$ $\Leftrightarrow$ $\exists\lambda \in (0, 1)$ such that, for each x^1 near x^0 and π near the origin, there is a minimizing sequence $\{x^k\}_{k\ge 1}$ to the function $x \mapsto \|f(x) - \pi\|$ satisfying (3.12). With fixed $\pi = 0$, this condition describes calmness of S at $(0, x^0)$.*

Proof If ALG1 is applicable then convergence of $\{x^k\}$ and (3.2) yield with $p = f(x)$, since f is locally Lipschitz,

$$-Cd(x^{k+1},x^k) \le \|f(x^{k+1})-\pi\|-\|f(x^k)-\pi\| \le -\lambda\|f(x^k)-\pi\|$$

for some $C > 0$, hence (3.6) is now necessarily satisfied. The latter implies, up to a new constant in (3.2)(ii), that (3.2) and the requirements

$$\|f(x^{k+1})-\pi\|-\|f(x^k)-\pi\| \le -\lambda\, d(x^{k+1},x^k) \text{ and } d(x^{k+1},x^k) \ge c\,\|f(x^k)-\pi\|$$

(for $\lambda, c > 0$) are equivalent. Setting $\lambda := \min\{\lambda,\ c\}$, we need one constant only which gives (3.12). □

Remark 4 As in Corollary 1, in order to show the related stability, it suffices to verify that (3.12) holds for x^1 near x^0 and appropriate x^2 only. Moreover, due to Remark 3(iii), Corollary 2 remains true after adding the requirement $f(x^k) \in \operatorname{conv}\{f(x^1), \pi\}$.

Calmness and the relative slack for inequality systems

In particular, Corollary 2 applies to system (2.5) after defining f by $f(x) = (g(x)^+, h(x))$.

However, for the sake of simplicity we assume that the equations are written as inequalities, and study, first with $I = \{1, \dots, m\}$, calmness of

$$\Sigma(p) = \{x \in X \mid g_i(x) \le p_i,\ \forall\, i \in I\} \tag{3.13}$$

at $(0, x^0)$ with locally Lipschitzian g_i and a Banach space X.

We write $g^m(x) = \max_i \; g_i(x)$ and define, for $g^m(x) > 0$, some *relative slack of* g_i in comparison with g^m,

$$s_i(x) = \frac{g^m(x) - g_i(x)}{g^m(x)} \quad (\geq 0). \tag{3.14}$$

In the special case of $g \in C^1, \; X = \mathbb{R}^n$, the following condition (3.16) differs just by the additionally appearing quantities $s_i(x)$ from the MFCQ-condition (or the Aubin property, cf. Lemma 1) for inequalities.

Theorem 3 *Let* $g^m(x^0) = 0$. *Then* Σ (3.13) *is calm at* $(0, x^0)$ *if and only if there exist some* $\lambda \in (0, 1)$ *and a neighborhood* Ω *of* x^0 *such that the following holds:*
For all $x \in \Omega$ *with* $g^m(x) > 0$ *there exist* $u \in \operatorname{bd} B$ *and* $t > 0$ *satisfying*

$$\frac{g_i(x+tu) - g_i(x)}{t} \leq \frac{g^m(x) - g_i(x)}{t} - \lambda \; \forall i \;\; \textit{and} \;\; \lambda g^m(x) \leq t \leq \frac{1}{\lambda} g^m(x). \tag{3.15}$$

Moreover, if $g \in C^1$, *one may delete* t *and replace* (3.15) *by*

$$Dg_i(x^0)u \;\leq\; \frac{s_i(x)}{\lambda} - \lambda \quad \forall i. \tag{3.16}$$

Proof We study the system $f(x) := (g^m)^+(x) = r$ which is calm at $(0, x^0)$ iff so is Σ. In accordance with Remark 4, calmness means that some $\lambda \in (0, 1)$ satisfies:
$\forall x$ near x^0 with $g^m(x) > 0 \; \exists x'$ such that

$$(g^m)^+(x') - g^m(x) \leq -\lambda \, d(x', x) \;\; \text{and} \;\; d(x', x) \geq \lambda \, g^m(x). \tag{3.17}$$

Defining $Q_i = \frac{g_i(x') - g_i(x)}{d(x',x)}$ we have $g_i(x') = g_i(x) + d(x', x)Q_i$. Then the first condition of (3.17) implies

$$\begin{array}{ll} d(x', x) & \leq \frac{g^m(x)}{\lambda} \;\; \text{and} \\ g_i(x) + d(x', x)Q_i \quad (= g_i(x')\,) & \leq g^m(x) - \lambda d(x', x) \; \forall \, i \end{array} \tag{3.18}$$

and vice versa.Writing here $x' = x + tu$ with $\|u\| = 1$ and $t > 0$, so (3.17) claims exactly (3.15). It remains to investigate the case of $g \in C^1$. First note that (3.15) yields, due to $\lambda g^m(x) \leq t$,

$$\frac{g_i(x+tu) - g_i(x)}{t} \leq \frac{g^m(x) - g_i(x)}{\lambda \; g^m(x)} - \lambda \; \forall i \;\; \text{and} \;\; \lambda g^m(x) \leq t \leq \frac{1}{\lambda} g^m(x). \tag{3.19}$$

Since also uniform convergence

$$\sup_{i \in I, \; \|u\|=1} \left| \frac{g_i(x+tu) - g_i(x)}{t} - Dg_i(x^0)u \right| \to 0 \quad \text{as } x \to x^0, t \downarrow 0 \tag{3.20}$$

is valid, now (3.19) implies (3.16) (with possibly smaller λ). Hence (3.15) implies (3.16). Conversely, having (3.16) it suffices to put $t = \lambda g^m(x)$ in order to obtain (3.15) (possibly with smaller λ, too). This completes the proof. □

Notes (modifying Theorem 3):

(i) Instead of considering all $x \in \Omega$ with $g^m(x) > 0$, it suffices to regard only

$$x \in \Omega \;\text{ with }\; 0 < g^m(x) < \lambda \|x - x^0\| \tag{3.21}$$

since, for $g^m(x) \geq \lambda \|x - x^0\|$, it holds the trivial calmness estimate

$$\text{dist}\,(x, \Sigma(0)) \leq \|x - x^0\| \leq \frac{1}{\lambda} g^m(x) \tag{3.22}$$

and one may put $u = \frac{x^0 - x}{\|x^0 - x\|}$, $t = \|x^0 - x\|$ in the theorem. Since λ may be arbitrarily small, so *calmness depends only on sequences* $x \to x^0$ *satisfying* $g^m(x) = o(x - x^0) > 0$.

(ii) Trivially, (3.15) is equivalent to

$$g^m(x + tu) \leq\; g^m(x) - \lambda t \quad \text{and} \quad \lambda^2 g^m(x) \leq \lambda t \leq\; g^m(x). \tag{3.23}$$

(iii) For $g \in C^1$, condition (3.16) can be replaced by

$$Dg_i(x)u \;\leq\; \frac{s_i(x)}{\lambda} - \lambda \quad \forall i \tag{3.24}$$

(possibly with smaller Ω and λ). Moreover, if $s_i(x) \;\geq\; \sqrt{\lambda}$, i.e., $(1 - \sqrt{\lambda})g^m(x) \;\geq\; g_i(x)$, and λ is small enough, then (3.16) (and (3.24)) is always satisfied. Hence, recalling (3.21) and (3.22), *only points x near x^0 with* $\text{dist}\,(x, \Sigma(0)) > \lambda^{-1} g^m(x)$ *and* (3.21) *and constraints g_i with $g_i(x) > (1 - \sqrt{\lambda})\, g^m(x)$ are of interest for condition* (3.16).

The torus-condition (3.15)*:*

Generally, since the stepsize t in condition (3.15) is restricted to a compact interval in the positive half-line, the left-hand side in (3.15) compares points the difference tu of which belongs to a torus. Therefore, without additional assumptions, the assigned quotients cannot be described by known (generalized) derivatives since such derivatives consider always *arbitrarily close* preimage points. The quotients on the right-hand side

$$\frac{g^m(x) - g_i(x)}{t} = \frac{g^m(x)}{t} s_i(x) \quad \text{where} \quad \frac{g^m(x)}{t} \in [\lambda, \; \frac{1}{\lambda}]$$

may vanish or not as $x \to x^0$.

Remark 5 (*Infinitely many constraints.*) As in usual semi-infinite programs, one can consider Σ (3.13) with a compact topological space I, $\|p\| = \sup_i |p_i|$, and a continuous map $(i, x) \mapsto g_i(x)$ which is uniformly (in view of $i \in I$) locally Lipschitz w.r. to x near x^0. Further, write $g \in C^1$ if all $Dg_i(x)$ w.r. to x exist and are continuous on $I \times X$. Then, due to (3.20),

> *Theorem 3 and the related Notes remain true without changing the proof. The same holds for all subsequent statements of this subsection, in particular for Theorem 4.*

Using the relative slack for deforming and solving system $g(x) \le 0$, $g \in C^1$

In the C^1 case, the above calmness condition for Σ (3.13) becomes stronger after adding $\varepsilon\|x - x^0\|^2$ to all g_i: Indeed, the set of all $x \in \Omega$ with

$$g^m(x) + \varepsilon\|x - x^0\|^2 > 0$$

is not smaller than before and the relative slack s_i is now smaller. Hence, the original system is calm whenever so is the perturbed one.

In order to *solve* the inequality system $\Sigma(0)$ of (3.13), we recall that the minimizing sequence of Corollary 2 can be obtained by the successive assignment $x \mapsto x' = x + tu$, cf. (3.8).

It is clear that finding u may be a hard task in general. However, if $g \in C^1$, we may replace (3.16) by condition (3.24) and put $t = \lambda g^m(x)$. This yields both an algorithm for finding some $x_\pi \in \Sigma(0)$ and a calmness criterion as well.

ALG2: Given $x^k \in X$ and $\lambda_k > 0$, solve the system

$$Dg_i(x^k)u \le \frac{s_i(x^k)}{\lambda_k} - \lambda_k \quad \forall i \quad \|u\| = 1. \tag{3.25}$$

Having a solution u, put $x^{k+1} = x^k + \lambda_k g^m(x^k)u, \ \lambda_{k+1} = \lambda_k$,
otherwise put $x^{k+1} = x^k, \quad \lambda_{k+1} = \frac{1}{2}\lambda_k$.

Corollary 3 (ALG2). *Let $g \in C^1$. Then Σ is calm at $(0, x^0)$ if and only if there is some $\alpha > 0$ such that, for $\|x^1 - x^0\|$ small enough and $\lambda_1 = 1$, it follows $\lambda_k \ge \alpha$ $\forall k$. In this case, the sequence x^k converges to some $x_\pi \in \Sigma(0)$ and*

$$g^m(x^{k+1}) \le (1 - \beta^2)g^m(x^k) \text{ whenever } 0 < \beta < \alpha \text{ and } g^m(x^k) > 0. \tag{3.26}$$

Proof The first statements follow from Corollary 2 and Theorem 3. The estimate is ensured by formula (3.23) and $t = \lambda g^m(x)$. □

We used condition $\|u\| = 1$ in (3.25) for obtaining the simple estimates (3.26). If one requires $\|u\| \le 1$ instead (*in order to define a more convenient convex auxiliary system*), then Corollary 3 is still true, only formula (3.26) becomes more complicated.

ALG3: Given $x^k \in X$ *and* $\lambda_k > 0$, solve the (convex) system

$$Dg_i(x^k)u \le \frac{s_i(x^k)}{\lambda_k} - \lambda_k \quad \forall i \ \|u\| \le 1. \tag{3.27}$$

Having a solution u, put $\quad x^{k+1} = x^k + \lambda_k g^m(x^k)u, \ \lambda_{k+1} = \lambda_k,$
otherwise put $\quad x^{k+1} = x^k, \qquad \lambda_{k+1} = \frac{1}{2}\lambda_k.$

Theorem 4 (ALG3). *Let $g \in C^1$. Then Σ is calm at $(0, x^0)$ if and only if there is some $\alpha > 0$ such that, for $\|x^1 - x^0\|$ small enough and $\lambda_1 = 1$, it follows $\lambda_k \ge \alpha$ $\forall k$. In this case, the sequence x^k converges to some $x_\pi \in \Sigma(0)$, and it holds*

$$g^m(x^{k+1}) \le (1 - \beta^2)g^m(x^k) \text{ whenever } 0 < \beta < \alpha^2/C \text{ and } g^m(x^k) > 0 \tag{3.28}$$

with $C = 1 + \sup_i \|Dg_i(x^0)\|$.

Proof We verify the first statement, the estimate then follows from the proof. In view of Corollary 3, we have only to show that $\lambda_k \ge \alpha > 0$ for ALG3 implies $\inf \lambda_k > 0$ for ALG2. Hence let $\lambda_k \ge \alpha > 0$ hold, with x^1 near x^0, for ALG3. We obtain $\|u\| > 0$ from (3.27) since there is always some $i = i(k)$ with $s_i(x^k) = 0$. Moreover, for x^k close to x^0, we have $\|Dg_{i(k)}(x^k)\| \le C$ and obtain even $\|u\| \ge \lambda_k/C$. Setting now

$$u' = u/\|u\| \text{ and } \lambda'_k = \lambda_k \|u\| \tag{3.29}$$

we generate the same points

$$x^{k+1} = x^k + \lambda_k g^m(x^k)u = x^k + \lambda'_k g^m(x^k)u', \tag{3.30}$$

and $\lambda_k \ge \alpha$ implies

$$\lambda'_k = \lambda_k \|u\| \ge \lambda_k^2/C \ge \alpha' := \alpha^2/C.$$

Finally, it holds for all i, as required in (3.25),

$$Dg_i(x^k)u' = \frac{Dg_i(x^k)u}{\|u\|} \le \frac{s_i(x^k)}{\lambda_k \|u\|} - \frac{\lambda_k}{\|u\|} = \frac{s_i(x^k)}{\lambda'_k} - \frac{\lambda'_k}{\|u\|^2} \le \frac{s_i(x^k)}{\lambda'_k} - \lambda'_k.$$

This tells us that, up to getting new constants, it suffices to claim $\|u\| \le 1$ in ALG2. The estimate (3.26) implies, due to (3.29) and (3.30),

$$g^m(x^{k+1}) \le (1 - \beta^2)g^m(x^k) \text{ whenever } 0 < \beta < \alpha' \text{ and } g^m(x^k) > 0.$$

This is exactly (3.28). □

In order to demonstrate the content of system (3.27) for different original problems we consider two examples.

Example 1. Ordinary differential equation:

Let $X = C[0, 1]$ consist of functions $x = x(t)$ and identify $I = [0, 1]$ and $i = t$ in order to describe constraints $g_t(x) := g(x(t)) \le 0 \; \forall t$.

Put $G(x) = x - y$ with

$$y(t) = a + \int_0^t f(x(s), s)ds, \quad 0 \le t \le 1, \; f \in C^1.$$

Then $G(x) = 0$ describes the solutions of $\dot{x} = f(x, t), \; x(0) = a$. With

$$g_t(x) = G(x)(t) = x(t) - a - \int_0^t f(x(s), s)\, ds$$

the differential equation becomes $g_t(x) \le 0, -g_t(x) \le 0 \; \forall t$. Further, it holds

$$DG(x)(u)(t) = u(t) - \int_0^t f_x(x(s), s)\, u(s)\, ds$$

and the inequalities (3.27) require with

$$m_k = \sup_t |g_t(x^k)|, A_k(s) = f_x(x^k(s), s) \text{ and } \|u\| \le 1$$

for all t,

$$\begin{aligned} u(t) - \int_0^t A_k(s)\, u(s)\, ds \; &\le \frac{s_t(x^k)}{\lambda_k} - \lambda_k, \quad \text{where } s_t(x^k) = \frac{m_k - g_t(x^k)}{m_k}, \\ -u(t) + \int_0^t A_k(s)\, u(s)\, ds \; &\le \frac{s_t(x^k)}{\lambda_k} - \lambda_k, \quad \text{where } s_t(x^k) = \frac{m_k + g_t(x^k)}{m_k}. \end{aligned}$$

The auxiliary problems of ALG3 are now linear integral inclusions (which can be solved via discretization arbitrarily precise) and $x^{k+1} = x^k + m_k \lambda_k u$.

Example 2. Stationary points for optimization in $\mathbb{R}^n$: For the problem

$$\min f_0(x) \quad s.t.\, x \in \mathbb{R}^n, \; f_\mu(x) \le 0, \quad f_0, f_\mu \in C^2, \; \mu = 1, \ldots, m \tag{3.31}$$

the Karush-Kuhn-Tucker (KKT) points $(x, y) \in \mathbb{R}^{n+m}$ are given by $g(x, y) \le 0$ where

$$\begin{aligned} &g_j^1(x, y) = \tfrac{\partial f_0(x)}{\partial x_j} + \textstyle\sum_\mu y_\mu \tfrac{\partial f_\mu(x)}{\partial x_j}, \qquad g_j^2(x, y) = -g_j^1(x, y), \\ &g_\mu^3(x, y) = f_\mu(x), \quad g_\mu^4(x, y) = -y_\mu, \; g_\mu^5(x, y) = -y_\mu f_\mu(x). \end{aligned} \tag{3.32}$$

Clearly, $g^m(x, y)$ denotes the maximum of all functions, and the first set of conditions in (3.27) requires that, for $(u, v) \in \mathbb{R}^{n+m}$, $\|(u, v)\| \leq 1$ and

$$s_j^1(x^k, y^k) := \frac{g^m(x^k, y^k) - g_j^1(x^k, y^k)}{g^m(x^k, y^k)} \;,$$

one has

$$D_x g_j^1(x^k, y^k)u + D_y g_j^1(x^k, y^k)v \;\leq\; \frac{s_j^1(x^k, y^k)}{\lambda_k} - \lambda_k \,. \tag{3.33}$$

Analogously, the other conditions of ALG3 are defined by linear inequalities. With

$$s_\mu^5(x^k, y^k) := \frac{g^m(x^k, y^k) - g_\mu^5(x^k, y^k)}{g^m(x^k, y^k)} \;,$$

we consider the last ones explicitly

$$-\,(\; y_\mu^k D_x f_\mu(x^k)u + f_\mu(x^k)v_\mu \;) \;\leq\; \frac{s_\mu^5(x^k, y^k)}{\lambda_k} - \lambda_k \tag{3.34}$$

in order to check the role of strict complementarity in the KKT system restricted to the set $\Delta := \{(x, y) \mid \max\{-y_\mu, f_\mu(x)\} \geq 0 \, \forall \mu \in I^0\}$, where $I^0 := \{\mu \mid y_\mu^0 = f_\mu(x^0) = 0\}$. Note that ALG3 can easily be adapted to this case.

If strict complementarity is violated at the reference point, i.e., $y_\mu^0 = f_\mu(x^0) = 0$ (for some μ), then $g_\mu^5(x^k, y^k) > 0$ yields, for

$\{(x^k, y^k)\} \subset \Delta$ with $(x^k, y^k) \to (x^0, y^0)$,

$$g_\mu^5(x^k, y^k) << \max\{-y_\mu^k,\; f_\mu(x^k)\} \leq g^m(x^k, y^k) \;\text{ and }\; s_\mu^5(x^k, y^k) \;\to 1.$$

If $g_\mu^5(x^k, y^k) \leq 0$ then even $s_\mu^5(x^k, y^k) \geq 1$ follows. Hence condition μ of (3.34) is always fulfilled for $\{(x^k, y^k)\} \subset \Delta$ with sufficiently small $\|(x^k, y^k) - (x^0, y^0)\| + \lambda_k$. This implies for problem (3.31):

Corollary 4 *Calmness of the KKT-system restricted to Δ does not depend on the fact whether strict complementarity is violated or not at the reference point.*

The same is obviously true if additional equations are required in (3.31), and remains also valid for related local descriptions of variational inequalities where $Df_0(x)$, in (3.32), is replaced by a C^1- vector function $H = H(x) \in \mathbb{R}^n$.

4 Interrelations with known methods

Next we consider particular methods and their relations to ALG1 and PRO(γ), respectively.

4.1 PRO(γ) as Feijer and Penalty method

Interpretations of PRO(γ) are possible in the form of classical first-order methods.

PRO(γ) *as Feijer method:*

The construction of the sequence can be understood as a Feijer method with respect to the norm $\|.\|_X + \frac{1}{\lambda}\|.\|_P$ and the subsets $M_1 = \{\pi\} \times X$ and $M_2 = \text{gph}\, S$ of (P, X):
Given $z^k = (p^k, x^k)$,
find first the point $U^k = (\pi, x^k)$ by projection of z^k onto M_1 (a trivial step) and next some V^k by projection of U^k onto M_2 (up to error $\gamma \|p^k - \pi\|$).
Write $z^{k+1} = V^k = (p^{k+1}, x^{k+1})$ and repeat.

PRO(γ) *as penalty method:*

The term $\frac{1}{\lambda}\|p' - \pi\|$ in the objective of (3.9) can be understood as penalizing the requirement $p' = \pi$. So we simply solve (again approximately)

$$\min\ d(x', x^k) \quad s.t.\ (p', x') \in \text{gph}\, S,\ \ p' = \pi$$

by penalization of the equation $p' = \pi$. The quantity p^k then turns out to be the current approximation of $p' = \pi$, assigned to x^k.
Condition (3.10) requires, as already (3.2), *linear convergence*. Hence, summarizing, this ensures

Corollary 5 *Calmness and the Aubin property at $z^0 = (p^0, x^0)$ are equivalent to uniform locally linear convergence of the classical solution methods mentioned above with initial points x, p, π near z^0, where*

- (i) *one has to require $\pi = p^0$ (calm) and π near p^0 (Aubin property), respectively, and*
- (ii) *approximate solutions (up to error $\gamma \|p^k - \pi\|$) may be permitted.*

For solving the auxiliary problems, several approaches are thinkable in concrete situations.
So, for a continuous function $f : X \to P$, $S = f^{-1}$ and $\pi = 0$, PRO(γ) requires to put $p^k = f(x^k)$ and to find x^{k+1} such that

$$\lambda d(x^{k+1}, x^k) + \|f(x^{k+1})\| \le \lambda\gamma \|f(x^k)\| + \inf_{x'}[\lambda d(x', x^k) + \|f(x')\|]. \tag{4.1}$$

The crucial condition (3.10) means explicitly

$$\lambda\, d(x^{k+1}, x^k) \ +\ \|f(x^{k+1})\| \ \le\ \theta\ \|f(x^k)\| \quad \text{with some } \theta < 1 \tag{4.2}$$

and is, for a locally Lipschitz function f, equivalent to (3.2) and (3.12) (up to choice of the constants), cf. Corollary 2.

Using Ekeland's variational principle [12] and continuity of $\|f\|$, each x^{k+1} can be obtained by (exact) minimizing

$$\lambda\, d(x', x^k) + \|f(x')\|, \quad x' \in X,$$

i.e., by computing "Ekeland points" with weight λ (which always exist). For a Hilbert space X, this is a *"proximal point" step* applied to $\|f\|$ where $\|x' - x^k\|^2$ is replaced by $\lambda\|x' - x^k\|$.

4.2 The situation for the Lyusternik/Graves theorem I

To show how condition (3.10) can be verified via Newton's method, we consider PRO(γ) under the assumption of the closely connected (cf. [19]) classical theorems of Lyusternik [28] and Graves [15]:

$$\text{Let } F = g \text{ be a } C^1 \text{ function and } DF(x^0) \text{ map } X \text{ onto } P. \tag{4.3}$$

Then we have $(p^k, x^k) = (F(x^k), x^k),\ F(x') = F(x^k) + DF(x^k)(x' - x^k) + o(x' - x^k)$ and PRO(γ) requires to minimize (approximately)

$$d(x', x^k) + \frac{1}{\lambda}\|F(x') - \pi\|.$$

For this reason, forget $d(x', x^k)$ and consider approximate Newton equations to $F(x') = \pi$, namely

$$p^k - \pi + A_k(x' - x^k) = 0 \tag{4.4}$$

with some linear operator A_k. Since $DF(x^0)$ maps onto P there are positive c, C such that, if $\|A_k - DF(x^0)\| < c$, there exists a solution satisfying

$$d(x', x^k) \le C\|p^k - \pi\|. \tag{4.5}$$

Clearly, usually one takes $A_k = DF(x^k)$ and chooses x', among the solutions to (4.4) sufficiently close to x^k. This corresponds to the distance rule (4.6) in the context of successive approximation.

Now specify $\lambda \in (0,\ C^{-1})$, choose $\delta > 0$ with $\theta := (\delta + \lambda)C < 1$ and put $(p^{k+1}, x^{k+1}) = (F(x'), x')$. Next we apply standard arguments: Since

$$\begin{aligned} o(x' - x^k) &:= F(x') - F(x^k) - DF(x^k)(x' - x^k) \\ &= \int_0^1 [DF(x^k + t(x' - x^k)) - DF(x^k)](x' - x^k)dt \end{aligned}$$

and

$$\|DF(x^k + t(x' - x^k)) - DF(x^k)\| < \frac{\delta}{2}$$

hold for all x', x^k, sufficiently close to x^0, it follows

$$\|o(x' - x^k)\| \le \frac{\delta}{2}\, d(x', x^k).$$

Using (4.4) and also $\|DF(x^k) - A_k\| < \frac{\delta}{2}$ (otherwise decrease c), this yields

$$\begin{aligned}\|p^{k+1} - \pi\| &= \|F(x^k) + DF(x^k)(x' - x^k) + o(x' - x^k) - \pi\| \\ &= \|(DF(x^k) - A_k)(x' - x^k) + o(x' - x^k)\| \\ &\le \delta\, d(x', x^k) \le \ \delta C\|p^k - \pi\|.\end{aligned}$$

Recalling (4.5) we thus obtain (3.10) due to

$$\lambda\, d(x^{k+1}, x^k) \ + \ \|p^{k+1} - \pi\| \ \le \lambda C\|p^k - \pi\| + \ \delta C\|p^k - \pi\| = \ \theta\|p^k - \pi\|,$$

and (3.10) implies (3.9) with $\gamma = \theta/\lambda$.

4.3 PRO(γ) for "contractive" multifunctions

Let $T : X \rightrightarrows X$ (closed) obey the Aubin property with rank $L = q < 1$ at (x^0, t^0) and let $d_0 := d(t^0, x^0)$ be small enough such that $\frac{d(t^0,x^0)}{1-q} < \hat{\varepsilon} := \min\{\varepsilon, \delta\}$ (with ε, δ from Def. 1). Then the existence of a fixed point $\hat{x} \in T(\hat{x})$ near x^0 can be shown by *modified successive approximation* based on the steps $x^1 := t^0$ and

$$\text{select } x^{k+1} \in T(x^k) \quad \text{with} \quad d(x^{k+1}, x^k) \le q\, d(x^k, x^{k-1}) \tag{4.6}$$

where $k \ge 1$, $\hat{x} = \lim x^k$ and $d(\hat{x}, x^0) \le \frac{d(t^0,x^0)}{1-q}$ since one obtains a Cauchy sequence with

$$d(x^{k+1}, x^0) \le d(x^{k+1}, x^k) + \cdots + d(x^1, x^0) \le \left(\sum_{n\ge 0} q^n\right) d(x^1, x^0). \tag{4.7}$$

Next we show how PRO(γ) can be used to derive the same result.

Put $G(x) = T(x) - x$ and apply PRO(γ) to $S = G^{-1}$ with $\pi = 0$. Hence, given $(p^k, x^k) \in \operatorname{gph} S,\ k \ge 1$ we have to find $(p^{k+1}, x^{k+1}) \in \operatorname{gph} S$ such that

$$d(x^{k+1}, x^k) + \frac{\|p^{k+1}\|}{\lambda} \le \inf_{(p',x') \in \operatorname{gph} S} [\, d(x', x^k) + \frac{\|p'\|}{\lambda})\,] \ + \gamma\|p^k\|. \tag{4.8}$$

By the structure of G it holds

$$p^k = t^k - x^k \in T(x^k) - x^k = G(x^k) \ \text{ and } \ p' = t' - x' \in T(x') - x' = G(x').$$

So (4.8) requires $t^{k+1} \in T(x^{k+1})$ and

$$d(x^{k+1}, x^k) + \frac{\|t^{k+1} - x^{k+1}\|}{\lambda} \le \inf_{(x',t') \in \operatorname{gph} T} [\, d(x', x^k) + \frac{\|t' - x'\|}{\lambda}\,] \ + \gamma\|p^k\|. \tag{4.9}$$

Here, $p^0 = t^0 - x^0$ with norm d_0 is given and initial points $(p^1, x^1) \in \operatorname{gph} S$ may be taken arbitrarily close to (p^0, x^0). Because of $p^1 = t^1 - x^1 \in T(x^1) - x^1$ the latter means that $\|x^1 - x^0\|$ and $\|t^1 - t^0\|$ are sufficiently small. In particular, one finds for $x^1 = t^0$ (and more general for small $\|x^1 - x^0\| + d_0$) some $t^1 \in T(x^1)$ with $\|t^1 - t^0\| \le q\|x^1 - x^0\|$.

Now, beginning with $k = 1$, points (p^k, x^k) and $t^k = p^k - x^k \in T(x^k)$ are given. Since λ is (generally) small, the main term $\|t^{k+1} - x^{k+1}\|$ has to be sufficiently small in (4.9). Using the Aubin property and $t^k \in T(x^k)$, this induces to put $x^{k+1} = t^k$ $(k \ge 1)$ and, as long as $(x^k, t^k) \in \operatorname{gph} T$ is close enough to (x^0, t^0), to select

$$t^{k+1} \in T(x^{k+1}) \quad \text{with} \quad d(t^{k+1}, t^k) \le q\, d(x^{k+1}, x^k). \tag{4.10}$$

So one obtains $\|p^{k+1}\| = \|t^{k+1} - x^{k+1}\| \le q\|t^k - x^k\| = q\|p^k\|$ and

$$\begin{aligned} \lambda d(x^{k+1}, x^k) + \|p^{k+1}\| &\le \lambda d(x^{k+1}, x^k) + q d(x^{k+1}, x^k) \\ &= (\lambda + q)\|t^k - x^k\| \;=\; (\lambda + q)\|p^k\|. \end{aligned}$$

Thus the crucial estimate (3.10) holds true if λ satisfies $\lambda + q = \theta < 1$. The inequalities

$$\|p^{k+1}\| = \|t^{k+1} - t^k\| \le q^k\|p^1\| \quad \text{and} \quad \|t^{k+1} - t^1\| \le \frac{q}{1-q}\,\|p^1\|$$

ensure that (t^k, x^k) remains in fact close to (t^0, x^0), provided that both

$$\|p^1\| = \|t^1 - x^1\| \;\text{ and }\; \|(p^1, x^1) - (p^0, x^0)\|$$

were small enough. As above, this can be guaranteed if (t^0, x^0) forms already a sufficiently exact "approximate fixed point" (with small d_0). Again, we obtain a feasible γ in (3.10) by setting $\gamma = \theta/\lambda = (\lambda + q)/\lambda$. The estimate (3.4) yields

$$d(\hat{x}, x^1) \le \frac{1}{\lambda}\|p^1\| = \frac{1}{\lambda}\|t^1 - t^0\| \le \frac{q}{\lambda}\|x^1 - x^0\| = \frac{q}{\theta - q}\|t^0 - x^0\|.$$

In consequence, the given straightforward realization (4.10) *of PRO(γ) coincides with successive approximation* (4.6) *if* $x^1 = t^0$.

5 Successive approximation and perturbed maps

Modified successive approximation is the typical method for proving the theorems of Lyusternik and Graves for functions $g : X \to P$ under the already mentioned assumptions (4.3). Similarly, it can be used for verifying the Aubin property (and computing related solutions) of multifunctions $\Gamma = (h + F)^{-1}$ after nonlinear perturbations h as in (1.2),

$$p \in h(x) + F(x),$$

where F is closed and $h : X \to P$ is locally a Lipschitz function, i.e.,

$$\|h(x') - h(x)\| \le \alpha\, d(x', x) \text{ for all } x', x \in x^0 + \delta_0 B \text{ and } \|h(x^0)\| \le \beta. \tag{5.1}$$

Then the key observations consist of two facts:

(i) There hold the identities

$$x \in \Gamma(p) \Leftrightarrow p - h(x) \in F(x) \Leftrightarrow x \in S(p - h(x)) \tag{5.2}$$

and (obviously) the composed mapping $T_p(x) = S(p - h(x))$ obeys the Aubin property with rank $q = L\alpha$ at $(x^0, p + h(x^0))$ if S obeys the Aubin property with rank L at $z^0 = (p^0, x^0)$.

(ii) If $T : X \rightrightarrows X$ obeys the Aubin property with rank $q < 1$ at $(x^1, x^2) \in \operatorname{gph} T$ and $d(x^2, x^1)$ is sufficiently small (compared with q and ε, δ in Def. 1), then modified successive approximation (4.6) can be applied with start at (x^1, x^2).

For $T = T_p$ and small $\|p - p^0\| + \alpha + \beta$ the existence of appropriate initial points x^1 and x^2 is ensured: With $x^1 = x^0$ (or x^1 close to x^0) there exists $x^2 \in T(x^1) = S(p - h(x^1))$ such that

$$d(x^2, x^0) \le L \,\|\, (p - h(x^1)) - p^0 \,\| \tag{5.3}$$

is arbitrarily small. So successive approximation can be really applied and fulfills again $x^k \to \hat{x} \in T(\hat{x})$ and

$$d(\hat{x}, x^1) \le \frac{d(x^2, x^1)}{1 - q}. \tag{5.4}$$

The inequalities (5.3), (5.4) can be directly used for proving the Aubin property of Γ and deriving estimates of solutions in terms of the perturbation of (1.2). For sharp dual estimates (using co-derivatives) we refer to [11].

Theorem 5 *Let $S = F^{-1}$ obey the Aubin property with rank L at z^0, let $\delta_0 > 0$ and let h satisfy (5.1). Then, if $\|\pi - p^0\|$, α and β are sufficiently small (depending on δ_0 and the constants L, ε, δ in Def. 1), the mapping $\Gamma = (h + F)^{-1}$ obeys the Aubin property at $(p^0 + h(x^0),\ x^0)$ with rank $\frac{L}{1-L\alpha}$ and, moreover, there exists some $x_\pi \in \Gamma(\pi)$ with*

$$d(x_\pi, x^0) \;\le\; \beta L + \frac{L}{1 - L\alpha}\,(\,\|\pi - p^0\| + \alpha\beta L\,). \tag{5.5}$$

Proof Since, for small $\|\pi - p^0\|$, α and β, the mapping $T_\pi = S(\pi - h(.))$ obeys the Aubin property near x^0 with rank $q = L\alpha < 1$, a proof can be directly based on (5.2), (5.3) and (5.4). Detailed estimates can be found in [23]. □

5.1 The situation for the Lyusternik/Graves theorem II

Under (4.3), let $A = Dg(x^0)$. Then $A^{-1} : P \rightrightarrows X$ is pseudo-Lipschitz by Banach's inverse mapping theorem applied to the canonical factorization $A : X|_{ker\ A} \to P$. Setting

$$F(x) = g(x^0) + A(x - x^0) \quad \text{and } h(x) = g(x) - F(x),$$

so the suppositions of Theorem 5 are satisfied for small δ_0. Since $\pi = h(x) + F(x) \Leftrightarrow \pi = g(x)$, this proves again local solvability of the latter equation with related estimates which is the Lyusternik/Graves theorem.

In order to *solve* $g(x) = \pi$ with initial point x^0, $p^0 = g(x^0)$ and π close to p^0, one may use that $h(x^0) = 0$. The iterations in (4.6), i.e., $x^{k+1} \in T(x^k)$, now stand for solving (with $k > 1$) the linear equation $F(x) = \pi - h(x^k)$, i.e.,

$$p^0 + Dg(x^0)(x - x^0) = \pi - h(x^k) \ \text{ and } \ d(x, x^k) \ \leq \ q\, d(x^k, x^{k-1}). \tag{5.6}$$

The equations of the projection method in (4.4) for $A_k \equiv Dg(x^0)$, namely

$$p^k - \pi + Dg(x^0)(x' - x^k) = 0$$

and (5.6) coincide after the equivalent settings

$$h(x^k) = p^k - p^0 - Dg(x^0)(x^k - x^0) \ \text{ and } \ p^k = h(x^k) + p^0 + Dg(x^0)(x^k - x^0). \tag{5.7}$$

This yields

Corollary 6 *The successive approximation steps* (5.6) *turn out to be approximate projection steps* (4.4)*, for* $A^k = Dg(x^0)$*, and vice versa, after the assignment* $p^k \leftrightarrow h(x^k)$ (5.7).

5.2 Modifying the inclusion in solution procedures

In the most applications of Theorem 5, the function h describes the difference between a C^1 function $g(x)$ and its local linearization $l_{x0}(x) := g(x^0) + Dg(x^0)(x - x^0)$ as in Sect. 5.1, where it's no matter whether the initial problem is an equation $g(x) = p$ or an inclusion $p \in F(x) := g(x) + G(x)$, cf. [33].

In view of solution methods, inclusions $p \in F(x)$ can be successfully replaced by $p \in h(x) + F(x)$ also in other situations, e.g. (Tykhonov regularization), if

$$h(x) = \varepsilon x \text{ and } F(x) = \partial f(x)$$

and $\partial f(x)$ is a subdifferential of a convex function f on a Hilbert space X.

In this context, it is worth to mention that, when applying the iterations (4.6) or Theorem 5, the mapping $T = T_\pi = S(\pi - h(.))$ can be changed by modifying h as long as α and β in (5.1) do not increase. So one may determine the sequence $\{x^k\}$ for functions h_k with vanishing constants α_k, β_k (5.1); hence also for $h_k = \varepsilon_k\, h_1$.

Then, $h_k + F$ can play the role of (or can be interpreted as) regularizations of F during the solution process.

However, adding h or h_k may also induce that the "equation" $0 \in F(x)$ will be solved by quite different methods. So, for a particular function h, after adding h or $-h$, the perturbations describe the application of a penalty and a barrier method, respectively, for determining critical points of optimization problems, cf. [21].

Estimates of the perturbed solutions to (1.2) (which do not depend on the sign of h in stability theory) then can be used in a unified way for both methods. For applications in the context of classical barrier methods under MFCQ, we refer to [16].

Acknowledgements The authors are indebted to two anonymous referees for their very detailed and constructive comments.

References

1. Aubin, J.-P., Ekeland, I.: Applied Nonlinear Analysis. Wiley, New York (1984)
2. Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: Non-Linear Parametric Optimization. Akademie-Verlag, Berlin (1982)
3. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
4. Burke, J.V.: Calmness and exact penalization. SIAM J. Control Optim. **29**, 493–497 (1991)
5. Burke, J.V., Deng, S.: Weak sharp minima revisited, Part III: Error bounds for differentiable convex inclusions. Math. Programm. Published online (2007)
6. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
7. Cominetti, R.: Metric regularity, tangent sets and second-order optimality conditions. Appl. Math. Optimiz. **21**, 265–287 (1990)
8. Dempe, S.: Foundations of Bilevel Programming. Kluwer, Dordrecht (2002)
9. Dolecki, S., Rolewicz, S.: Exact penalties for local minima. SIAM J. Control Optim. **17**, 596–606 (1979)
10. Dontchev, A.: Local convergence of the Newton method for generalized equations. Comptes Rendus de l'Acédemie Des Sciences de Paris **332**, 327–331 (1996)
11. Dontchev, A., Rockafellar, R.T.: Regularity and conditioning of solution mappings in variational analysis. Set-Valued Anal. **12**, 79–109 (2004)
12. Ekeland, I.: On the variational principle. J. Math. Anal. Appl. **47**, 324–353 (1974)
13. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementary Problems, Vol I and Vol II. Springer, New York (2003)
14. Fusek, P.: Isolated zeros of Lipschitzian metrically regular R^n functions. Optimization **49**, 425–446 (2001)
15. Graves, L.M.: Some mapping theorems. Duke Math. J. **17**, 11–114 (1950)
16. Grossmann, C., Klatte, D., Kummer, B.: Convergence of primal-dual solutions for the nonconvex log-barrier method without LICQ. Kybernetika **40**, 571–584 (2004)
17. Henrion, R., Outrata, J.: A subdifferential condition for calmness of multifunctions. J. Math. Anal. Appl. **258**, 110–130 (2001)
18. Henrion, R., Outrata, R.: Calmness of constraint systems with applications. Math. Program. Ser. B **104**, 437–464 (2005)
19. Ioffe, A.D.: Metric regularity and subdifferential calculus. Russ. Math. Surv. **55**, 501–558 (2000)
20. Klatte, D., Kummer, B.: Constrained minima and Lipschitzian penalties in metric spaces. SIAM J. Optim. **13**(2), 619–633 (2002)
21. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization—Regularity, Calculus, Methods and Applications. Kluwer, Dordrecht (2002)
22. Klatte, D., Kummer, B.: Strong Lipschitz stability of stationary solutions for nonlinear programs and variational inequalities. SIAM J. Optim. **16**, 96–119 (2005)
23. Klatte, D., Kummer, B.: Stability of inclusions: Characterizations via suitable Lipschitz functions and algorithms. Optimization **55**, 627–660 (2006)

24. Kummer, B.: Lipschitzian and pseudo-Lipschitzian inverse functions and applications to nonlinear programming. In: Fiacco, A.V. (ed.) Mathematical Programming with Data Perturbations, pp. 201–222. Marcel Dekker, New York (1998)
25. Kummer, B.: Metric regularity: characterizations, nonsmooth variations and successive approximation. Optimization **46**, 247–281 (1999)
26. Kummer, B.: Inverse functions of pseudo regular mappings and regularity conditions. Math. Program. Ser. B **88**, 313–339 (2000)
27. Li, W.: Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. SIAM J. Optim. **7**, 966–978 (1997)
28. Lyusternik, L.: Conditional extrema of functions. Math. Sbornik **41**, 390–401 (1934)
29. Mordukhovich, B.S.: Approximation Methods in Problems of Optimization and Control (in Russian). Nauka, Moscow (1988)
30. Outrata, J., Kočvara, M., Zowe, J.: Nonsmooth Approach to Optimization Problems with Equilibrium Constraints. Kluwer, Dordrecht (1998)
31. Robinson, S.M.: Stability theorems for systems of inequalities. Part II: Differentiable nonlinear systems. SIAM J. Numer. Anal. **13**, 497–513 (1976)
32. Robinson, S.M.: Generalized equations and their solutions, Part I: Basic theory. Math. Program. Study **10**, 128–141 (1979)
33. Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**, 43–62 (1980)
34. Robinson, S.M.: Generalized equations and their solutions Part II: Applications to nonlinear programming. Math. Program. Study **19**, 200–221 (1982)
35. Robinson, S.M.: Variational conditions with smooth constraints: structure and analysis. Math. Program. **97**, 245–265 (2003)
36. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)